



DESPERATELY SEEKING

Robert Filman • NASA Ames Research Center, Caelum Corp.
• filman@computer.org

It is far more impressive when others discover your good qualities without your help.

—Miss Manners (Judith Martin)

SOME ARE DISCOVERED, OTHERS FOUND OUT.

In earlier columns, we've looked at using search engines. Today, we start by grabbing the bull by the other horn, asking "How do I get search engines to notice my site?"

Search Engine Watch • <http://www.searchenginewatch.com>

Danny Sullivan

At Search Engine Watch, I found out more than I ever wanted to know about search engines. Pages include how to search, the histories of search engine companies, the economics of the search business, pointers to other pages discussing search engine issues, search engine trivia, performance reports, and, most germane to our question, discussions on how to get search engines to focus on a site.

You can explicitly submit your page to a search engine. However, the search engine may present your page as the 93rd best choice for its topic—even though we all know that your page contains all the greatest wisdom in the world on that topic and ought

to be listed first. Sullivan asserts the three keys to being highly ranked are

- having the searched-for keywords in the page title,
- having the searched-for keywords appear early on the page, and
- having the searched-for keywords appear frequently on the page.

Other factors include "link popularity" (lots of other pages point to this one), sites related to reviewed sites, and information in metatags, though use of these varies by engine. (Bribes also work in some cases.) Sullivan urges picking a two- or three-word "strategic key phrase" and including it in the title and early text. And, oh yes, make sure the page has relevant content.

Pitfalls include image maps (make sure you have real html links to your subpages), frames (some search engines can't handle them), and dynamically generated pages (for example, CGI). Sullivan advises submitting key pages to the major search engines and periodically checking that the pages remain referenced by those engines.

Sites perceived guilty of spamming (for example, by repeating a word too frequently, putting text in the same color as the background, or using an infinitesimal font) may be downgrad-

ed or ignored. Tricks with mixed results include

- having "bridge" (or "doorway" or "portal") pages, designed as luscious flies to webcrawlers, but which quickly propel human surfers to the real page;
- setting up a special portal page for a specific search engine and submitting that page just to that engine; and
- entering into "mutual referent" agreements with other sites, so your page appears to have many links to it.

If you've a particular search engine in mind, check <http://www.searchenginewatch.com/webmasters/features.html> for a description of the rating scheme and crawling behavior of all the major search engines.

The site also presents "EKG" metrics on search engines, graphing how deeply the major engines crawl over a site and how frequently they return to revalidate their links. Search Engine Watch gives good marks to AltaVista and Northern Light (comprehensive search and monthly refreshing). Inktomi was comprehensive, though not as fresh; Excite was not as deep, though revisited frequently. Infoseek had a middling comprehensiveness and refresh rating. Lycos scores poorly on depth, and Search Engine Watch didn't find a pattern in its refresh rate (though that might mean, for example, that Lycos refreshes the pages its users find useful, rather than being dumbly methodical).

Sullivan produces a free newsletter. I subscribed. For money, he promises even greater detail about search engine behavior. 🌿 🌿 🌿 🌿 🌿

Ask Jeeves • <http://www.ask.com>

Of course, all these search engines require you to properly phrase a query and sort the resulting wheat from chaff. How much better it would be if your questions were directly answered. At least that's the sales pitch of Ask Jeeves: Type in a natural language question, and it takes you directly to the relevant page—pages a human has picked out. The Spider paid a visit to try it out.

Asking Jeeves “Why is the sky blue?” produced the Blue Sky Exhibit at National Center for Atmospheric Research. I guess that’s a direct hit, but I never did see mention of lambda to the fourth. Asking Jeeves “What is the meaning of life?” produced <http://pobox.com/~sentience/tmol-faq/meaningoflife.html>, Frequently Asked Questions about the Meaning of Life, which assured me that “Nobody who tried to ‘find the meaning of life’ found it—what were you originally looking for?” and which stated its target audience was “people experiencing emotional distress because they don’t know why they’re getting up in the morning.”


Asking Jeeves “What are the 20 largest islands in the world?” got me straight to a list of the world’s 25 largest islands, but asking “What are the 20 largest islands in the United States?” brought me to the same page. (The largest island in the United States, Hawaii, at 10,433 sq. km., doesn’t begin to approach the size of global number 25, Sri Lanka, at 65,610 sq. km.) Jeeves also includes suggestions for similar questions.

My U.S. islands question prompted, “What are the islands of the world?” “Where can I find a concise encyclopedia article on islands?” “Where is/are the United States of America?” and “Where can I find a financial calculator to help me determine which is better: 15- or 30-year term mortgage?” I puzzled over that last one, until I realized that Jeeves was offering advice on how to finance the purchase of Sri Lanka.

Asking Jeeves “Where can I find the HTTP-NG proposed standard?” reduced Jeeves to consulting its conventional search engine friends: WebCrawler, Infoseek, Excite, and AltaVista. AltaVista returned nine relevant matches, WebCrawler and Excite, three each, while Infoseek remained hung up on the “Where can I find” part.

On the down side, Jeeves presents found pages within its own frame. This has the unfortunate side effect of not disclosing the URL of the found page. (Another justification for the good design advice: include a link to your home page on all the pages of

your site.) Good luck on finding your way back to something interesting.

Jeeves seems to use the Parry algorithm: Remember what people have asked you. Have good answers for those questions. (Get a human to find specific answers.) After a while, almost all the questions will repeat something you’ve already been asked. (Apparently more people care about sky color and life meaning than the area of Long Island or HTTP-NG.) Of course, *you* and *I* would never care about anything as unoriginal as something that had already interested someone else. So we don’t find Ask Jeeves all that useful. On the other hand, you can click a button on the Ask Jeeves page to see what everybody else is asking—a form of amusement sure to provide that reason for getting out of bed. 

WHO ARE YOU

Everyone keeps talking about LDAP, so I figured I’d have to find out what it is.

LDAP: Use as Directed •
<http://www.data.com/issue/990207/ldap.html>

Tim Howes, Netscape
LDAP stands for “Lightweight Directory Access Protocol,” with *directory* being used in the same sense as in “phone directory.” That is, a company can keep, say, the names and phone numbers (or names and public keys) of its employees in an LDAP directory, governed by a running server and accessible over a network. Then anyone that needs, say, an employee phone number, can remotely query

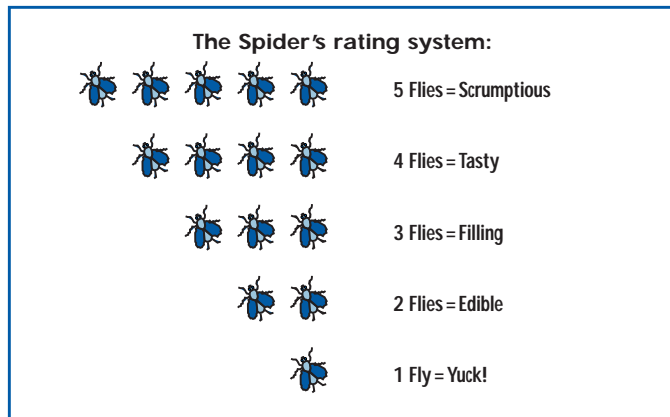
the LDAP server for that information. LDAP directories can be white pages, storing information about people, or yellow pages, storing information about resources. After all, supplying Joe’s phone number isn’t radically different from supplying the location of a virus protection program.

The best introductory discussion of the uses and limitations of LDAP I found was by Tim Howes, one of the authors of the original LDAP implementation at the University of Michigan.

While LDAP directories can do any pairwise association, Howe points out that you wouldn’t want to replace your relational database system, file system, or domain name server (DNS) with LDAP—it’s not engineered to do any of those tasks well. Rather, LDAP’s strength lies in ease of secure access.

LDAP arose as a simplified way to access X.500 services. ISO X.500 is a long-standing standard for a directory access system. Besides specifying a data model, X.500 has complex mechanisms for sharing data among directory servers and forwarding requests that can’t be handled locally. LDAP started out as a “lightweight” access to X.500 servers, one that could be run on a PC and didn’t require the full implementation of the ISO network protocol. LDAP has evolved to its own standalone system—LDAP servers are happy to run without an X.500 server in sight.

LDAP borrows the X.500 data model, which specifies a tree of domains (for example, organizations and their suborganizations, like DNS)



IEEE Internet Computing

ADVERTISER / PRODUCT INDEX

Advertiser / Products	Page Number
Copper Mountain	15
Cygnus Solutions	13
Eprise Corp.	15
FlowPoint	15
Hewlett Packard	13
IBM Corp.	13, 15
Informix	14
Infoseek	Cover 4
Larscom	15
Live Software	13
Lucent Technologies	15
Microsoft	13
Netopia	15
Network Appliance	13
Oracle	15
ParaSoft Corp.	1
Perforce Software	Cover 2
Platinum	15
Red Hat Software	13
Software Development Conf. '99	19, Cover 3
Sun Microsystems	13, 14

Boldface denotes advertisers in this issue.

Advertising Sales Offices

Sandy Aijala

10662 Los Vaqueros Circle, Los Alamitos, CA 90720-1314; Phone: (714) 821-8380; Fax: (714) 821-4010; saijala@computer.org

Kim Newman, Gene Selven

7291 Coronado Drive, Suite 8, San Jose, CA 95129; Phone: (408) 996-7401; Fax: (408) 996-7871; knewman@computer.org

For production information, conference, and classified advertising, contact Marian Anderson, *IEEE Internet Computing*, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720-1314; Phone: (714) 821-8380; Fax: (714) 821-4010; manderson@computer.org.

with the leaves representing particular individuals. Names are thus hierarchical. Leaves store a set of attribute value sequences—that is, Joe's phone numbers, address, fax number, favorite colors, and so forth. (See the diagram, "Getting to the Root of LDAP," at http://www.data.com/issue/990207/ldap_figure1.html.)

Howe identifies eight basic components of LDAP.

- The information model (attributes and values and the types of data). For example, the information model might posit an attribute "Address" of type "String."
- The LDAP schema, the collections of attributes making up the records to be stored on a server.
- The naming model, describing the hierarchy of names (for example, gov.nasa.arc.rfilman).
- The security model, including the ability to have both client and server mutually prove their identities.
- The functional model, the behaviors of an LDAP system, including adding, modifying, and deleting records; establishing connections; and searching.
- The LDAP protocol, defining the mapping of the functionality to TCP/IP.
- The application program interface, providing a standard set of calls for languages like C/C++, Java, and Perl to access a directory.
- The data interchange format, a simple text format for representing data entries.

Howe continues with a discussion of the limitations of the current LDAP standard and the directions for the next revision, version 3. Chief issues include improved replication and security. He closes with some examples of the benefits of using LDAP, such as maintaining a centralized repository of user passwords and phone information. 🌸 🌸 🌸 🌸

LDAP: the Next-Generation Directory • <http://www.sunworld.com/swol-10-1996/swol-10-ldap.html>

Dave Kosiur

Having gotten the LDAP overview from Howe, the Spider crawled over to SunWorld for some implementation detail. In this article, Dave Kosiur discusses some of the problems of directory systems and goes into some detail about X.500 data structures, LDAP replication strategies, and the differences between the protocols. He notes that X.500 is not a tabula rasa. Rather, it comes with 17 predefined classes and 40 basic attribute types particularly oriented towards "phone book" applications. Kosiur also gives examples of LDAP searches, including URLs that can be used to represent LDAP requests. Unfortunately, the article skips between factoids without presenting a unified thesis. 🌸 🌸

LDAP World •

<http://www.critical-angle.com/ldapworld/>
Innosoft

Want to find an LDAP product, a pointer to the LDAP FAQ, or track the changes in the LDAP standard? LDAP World is devoted to keeping you up-to-date. A good source of news and pointers. 🌸 🌸 🌸